A New Randomized Method for Solving Large Linear Systems

Elnur Gasanov¹ Vladislav Elsukov² Peter Richtárík¹,²

King Abdullah University of Science and Technology ¹ Moscow Institute of Physics and Technology ²

The Problem

Given $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, solve $Ax = b$.
That is, find $x \in \mathcal{L} \{ x \mid Ax = b \}$.

*Challenge:* This problem is difficult when $m \gg n$ (e.g., $m = 10^n$).

*Goal:* Design a new randomized method for finding an approximate solution quickly.

A New Method (NM)

- We propose a new method for solving the problem.
- The method does not require to calculate the Moore-Penrose pseudoinverse appearing in $H_k^\text{opt}$.

**Input:** Matrix $A$, vector $b$
**Parameters:** Same as BM, plus: carefully designed parameters $L_S > 0$ associated with every matrix $S$ for $k = 0, 1, 2, \ldots$

*Draw a fresh sample $S_k \sim \mathcal{D}$.*

$H_k^\text{opt} = \frac{1}{L_S}S(S^\top A B^\top A^\top S)^{-1}S_k^\top$

$x_{k+1} = x_k - \omega B^{-1}A^\top H_k^\text{opt}(Ax_k - b)$

**Output:** $x_k \approx x^*$

**Remarks:**
- $L_S$, is required to satisfy $L_S \geq \lambda_{\max}(S^\top A B^{-1}A^\top S)$
- The matrix $G^\text{opt} \equiv \mathbb{E}_{S \sim \mathcal{D}}[H_k^\text{opt}]$ controls the speed of the method

**NM vs SGD**

**Theorem [GER’19]** $\text{NM}$ is $\text{SGD}$ applied to the problem

$$\min_x f(x) \equiv \mathbb{E}[f_S(x)] \equiv \frac{1}{n} \|Ax - b\|_2^2$$

That is, $\text{NM}$ is equivalent to the method:

*Sample $S_k \sim \mathcal{D}$.*

$x_{k+1} = x_k - \omega \nabla f_S(x_k)$

Exactness

Define matrix $\Omega \equiv B^{-1}A^\top G^\text{opt} AB^{-1}$.

**Theorem [GER’19]** These statements are equivalent:

1. $\mathcal{L} = \mathcal{X} \{ \text{Argmin} f(x) \} \text{ ("exactness")}$
2. $\|G^\text{opt}\| \equiv \text{Null}(A)$
3. $\|G^\text{opt}\| \cap \text{Range}(A) = \emptyset$
4. $\Omega = \text{Null}(AB^{-1})$

Moreover, if $\mathcal{D}$ is absolutely continuous, then exactness for the $\text{NM}$ is equivalent to exactness for $\text{BM}$.

Convergence of iterates

**Theorem [GER’19]** Let $\Omega = U A U^\top$ be the eigenvalue decomposition. Then the random iterates generated by $\text{NM}$ converge linearly as follows:

$$\|x_k - x^*\|_B \leq \rho^k \|x_0 - x^*\|_B$$

where $\rho = \max \lambda_{\max}(1 - \omega \lambda_{\min}(\Omega))^2$.

Moreover, $\|x_k - x^*\|_B \leq (1 - \theta)^k \|x_0 - x^*\|_B$

where $\theta = \lambda_{\min}^+(\Omega)$.

Convergence of function values

Let $\lambda_{\min}$ and $\lambda_{\max}$ be the smallest positive and largest eigenvalues of $\Omega$, respectively. Choose $\omega \in \left[0, \frac{\lambda_{\max}}{\lambda_{\min}}\right]$. Then

$$\mathbb{E}[f(x_k)] \leq (1 - 2\lambda_{\min}^2 + \lambda_{\max}\omega^2)k f(x_0)$$

The optimal rate is achieved for $\omega = \lambda_{\min}/\lambda_{\max}$, in which case we get the bound

$$\mathbb{E}[f(x_k)] \leq (1 - \lambda_{\min}^2)^k f(x_0)$$

Comparison of rates

If exactness holds for both methods, then

$$\lambda_{\min}(\Omega) \leq \lambda_{\min}(B^{-1}A^\top G^\text{opt} AB^{-1})$$

If $B = I$, then

$$\lambda_{\min}(A^\top G_{\text{opt}} A) \leq \sup_s \lambda_{\min}(S^\top A A^\top S)$$

**Theorem [GER’19]** Let $B = I$ and assume the $i$th row of $A$ satisfies $\|A_i\|_1 = 1$ for all $i$. Further, let $\mathcal{D}$ be defined as follows: $S_i$ is a random column submatrix of $\mathcal{I}$ consisting of 2 columns. Finally, assume that $|\|A_i\|_1 - \alpha_0| < 1$ for all $i \neq j$.

Then

$$\lambda_{\min}(A^\top G_{\text{opt}} A) \leq \sqrt{1 + \varepsilon}$$

This means that $\text{NM}$ is at most $\sqrt{\frac{2}{\varepsilon}}$ times slower than $\text{BM}$ in terms of iterations. (However, it can be much faster in terms of cost of one iteration.)

Adaptive computation of $L_S$

If in each iteration $L_S$ satisfies

$$\|x_{k+1} - x_k\|_B \leq L_S \|x_k - x^*\|_B$$

then for $\text{NM}$ we have $\mathbb{E}[\|x_k - x^*\|_B] \to 0$ linearly.

References

